STT 223 Final Capstone

Due May 1, 2019 at 11:59pm

PUT YOUR NAME HERE

A small business owner is interested in examining the shopping habits of customers in order to get a better sense of inventory projections. The business owner examines customers on each of the seven days of the week to see if shopping habits differ by day. For each day, she collects two pieces of information: the number of items purchased, and the length of time customers spent shopping (in minutes). The data are as follows:

Day (j)	1	2	3	4	5	6	7
Items sold (y_j)							
Time (t_j)							

She has hired you to examine this data for interesting trends and to report back to her.

To begin your analysis, assume that the number of items purchased can be modeled by a Poisson distribution, with the parameter proportion to the length of time a customer shops:

$$y_j \sim Pois(\theta_j t_j)$$

Appropriate model priors to assign in this case are:

$$\begin{aligned} \theta_j|\alpha, \beta &\sim Gamma(\alpha, \beta) \\ \alpha &\sim Exp(\lambda = 1) \\ \beta &\sim Gamma(0.1, 1) \end{aligned}$$

1. Derive the joint posterior of $\theta_1, ..., \theta_7, \alpha, \beta|\mathbf{y}$ (done in class).
2. Derive the full conditional posterior distribution of the following (done is class):
    a. $\theta_j|\alpha, \beta, \mathbf{y}$
    b. $\alpha|\beta, \theta, \mathbf{y}$
    c. $\beta|\alpha, \theta, \mathbf{y}$

Recall the following: The Poisson distribution generally takes the following form:

$$p(y_j|\theta_j, t_j) = \frac{(\theta_j t_j)^{y_j} e^{-\theta_j t_j}}{y_j!}$$

The gamma distribution generally takes the following form:

$$p(\theta_j|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_j^{\alpha-1} e^{-\beta\theta_j}$$

The exponential distribution generally takes the following form:

$$p(\alpha|\lambda) = \lambda e^{-\alpha\lambda}$$

This ends the in-class portion of the capstone.

You should find that the full conditional posterior distribuitons are as follows:

$$\theta_j | \alpha, \beta, \mathbf{y} \sim Gamma(y_j + \alpha, t_j + \beta)$$

$$\beta | \alpha, \theta, \mathbf{y} \sim Gamma\left(J\alpha + 0.1, 1 + \sum_{j=1}^{J} \theta_j\right)$$

$$\alpha | \beta, \theta, \mathbf{y} \propto e^{-a} \prod_{j=1}^{J} \frac{\theta_j^{\alpha-1} \beta^{\alpha}}{\Gamma(\alpha)}$$

3. Use a Metropolis Hastings-within-Gibbs algorithm to take 10000 samples of each parameter from its full conditional posterior distribution. In particular, for each iteration, you will draw one sample of each θ_j from their respective gamma posterior distributions, one sample of β from its gamma posterior distribution, and one sample of α from its posterior distribution using a Metropolis-Hastings algorithm with a truncated normal proposal distribution centered at the current value of α and a proposal variance of 4. You can use starting values of each $\theta_j = 0.1$, $\beta = 1$, and $\alpha = 1$. Print the posterior mean for each parameter.

4. Construct and examine traceplots for each of the 9 parameters, and discuss whether you believe convergence has been met.

5. Now that you have results from one chain, repeat question 3 four more times using these starting values:

   a. $\theta_j = 5, \alpha = 0.5, \beta = 0.5$
   b. $\theta_j = 1, \alpha = 10, \beta = 10$
   c. $\theta_j = 0.5, \alpha = 1, \beta = 1$
   d. $\theta_j = 0.1, \alpha = 0.1, \beta = 0.1$

6. Considering the samples from all five chains, assess the convergence of each parameter using the Gelman-Rubin diagnostic.

7. Report the posterior mean and 95% posterior interval for each variable, using the combined results from all five chains. You may either put the results in the table below, rounding to three decimal places, or otherwise neatly present the results.

Parameter	Posterior Mean	Posterior Interval
θ_1		(,)
θ_2		(,)
θ_3		(,)
θ_4		(,)
θ_5		(,)
θ_6		(,)
θ_7		(,)
α		(,)
β		(,)

8. Nicely summarize in terms that the small business owner (who has a very limited statistics background) will understand what you did, what results you arrived at, and what conclusions you have for her. This write-up should be a short paragraph or two and provide a robust discussion of what you did in laypersons terms.